Condition of Lake Erie Largemouth Bass Sampled in the ODOW Nearshore Community Survey 2013-2017

Alex Benecke

January 4, 2018

Data Prep

```
Stock <- read.csv("Data/Clean-Data/largemouth-bass_Wr_Stock.csv") %>%
  filter(Year < 2017) %>%
  filterD(!is.na(Wr)) %>%
  arrange(Year,gcat)


Stock$fyr <- factor(Stock$fyr)
Stock$Year <- factor(Stock$Year)
```

```
headtail(Stock)
```

```
##     fyr Year Site FID Weight        Ws        Wr Length lcat20      gcat
## 1    13 2013   10   9    807  874.1340  92.31995    387    380 preferred
## 2    13 2013   11   1    968  934.6786 103.56501    395    380 preferred
## 3    13 2013   11   8   1159 1196.8993  96.83354    426    420 preferred
## 408  16 2016   18   8    479  336.1388 142.50066    289    280     stock
## 409  16 2016   18  14    466  351.6072 132.53427    293    280     stock
## 410  16 2016   18  23    473  375.7265 125.88945    299    280     stock
##     Age SexCon Sex
## 1     4      8   2
## 2     3      3   1
## 3     3      3   1
## 408   2      3   1
## 409   2      3   1
## 410   2      8   2
```

```
str(Stock)
```

```
## 'data.frame':    410 obs. of  13 variables:
##  $ fyr   : Factor w/ 4 levels "13","14","15",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Year  : Factor w/ 4 levels "2013","2014",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Site  : int  10 11 11 15 15 15 15 15 18 18 ...
##  $ FID   : int  9 1 8 167 NA 170 176 180 142 141 ...
##  $ Weight: int  807 968 1159 1144 927 982 1015 1000 942 941 ...
##  $ Ws    : num  874 935 1197 859 882 ...
##  $ Wr    : num  92.3 103.6 96.8 133.1 105.2 ...
##  $ Length: int  387 395 426 385 388 401 406 411 381 382 ...
##  $ lcat20: int  380 380 420 380 380 400 400 400 380 380 ...
##  $ gcat  : Factor w/ 3 levels "preferred","quality",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Age   : int  4 3 3 3 NA 4 5 5 3 3 ...
##  $ SexCon: int  8 3 3 3 3 8 8 8 8 3 ...
##  $ Sex   : int  2 1 1 1 1 2 2 2 2 1 ...
```

```
unique(Stock$fyr)
```

```
## [1] 13 14 15 16
## Levels: 13 14 15 16
```

Note

I removed the years 2012 and 2017. 2012 because only large fish have weight length data and more and differenct sites were samples. I removed 2017 due to differences in the survey.

Note

I am removing a fish from site 15 year 2013 because it appears to be a outlier (Wr > 200). Probably due to data entry error. **I went back and did this in 'Create-Wr-Gabelhouse-Data.Rmd' where I make the data file I use for this analysis.**

Summarize Relative Weight by Year

```
(Wr.Stock <- Summarize(Wr ~ Year, data = Stock) %>% arrange(Year))
```

```
##   Year   n     mean       sd   min     Q1 median    Q3   max
## 1 2013  97 113.3837 13.28398 76.14 104.40  113.0 122.2 143.8
## 2 2014 140 109.6148 15.74476 80.40  98.78  106.9 117.8 151.3
## 3 2015  67 109.8376 15.57996 78.22  98.53  108.6 120.8 149.8
## 4 2016 106 115.4942 13.83934 61.76 108.00  115.5 124.7 146.2
```

Note

The relative weight data contains only stock length individuals. This is so that I can easily compare the relative weight of fish with PSD. This is done despite the min TL being 150 mm. I may want to summarize relative weight for 150mm and greater length individuals in the future to see if young/small fish drive down or increase Wr.

Lets start exploring the relative weight data. I have two questions I would like to know the answer to.

1) does Wr differ among years?
2) does Wr differ among gabelhouse length categories?

First Lets see if Wr is different between years.

```
aov1 <- lm(Wr ~ Year, data = Stock)
# save(aov1,file = 'model-output/aov1.rda')
Anova(aov1)
```

```
## Anova Table (Type II tests)
##
## Response: Wr
##           Sum Sq  Df F value   Pr(>F)
## Year        2588   3  4.0016 0.007933 **
## Residuals  87529 406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mc1 <- glht(aov1, mcp(Year = "Tukey"))
summary(mc1)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ Year, data = Stock)
##
## Linear Hypotheses:
##                   Estimate Std. Error t value Pr(>|t|)
## 2014 - 2013 == 0  -3.7690     1.9397  -1.943   0.2102
## 2015 - 2013 == 0  -3.5461     2.3324  -1.520   0.4242
## 2016 - 2013 == 0   2.1104     2.0631   1.023   0.7346
## 2015 - 2014 == 0   0.2228     2.1812   0.102   0.9996
## 2016 - 2014 == 0   5.8794     1.8904   3.110   0.0108 *
## 2016 - 2015 == 0   5.6566     2.2916   2.468   0.0658 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Looks like Wr *is significantly different* between years (One-Way ANOVA, $F_{3,406} = 4.00$, p = 0.007933). There is *no significant difference* in relative weight between 2013 and 2014 (Tukey HSD, t = -1.94, p = 0.2101), 2013 and 2015 (Tukey HSD, t = -1.52, p = 0.4241), 2013 and 2016 (Tukey HSD, t = 1.02, p = 0.7346), 2015 and 2014 (Tukey HSD, t = 0.10, p = 0.9996), and 2015 and 2016 (Tukey HSD, t = 2.47, p = 0.0656). However, relative weight *is significantly different* between 2014 and 2016 (Tukey HSD, t = 3.11, p = 0.0107).

constructing a plot of Wr and Year

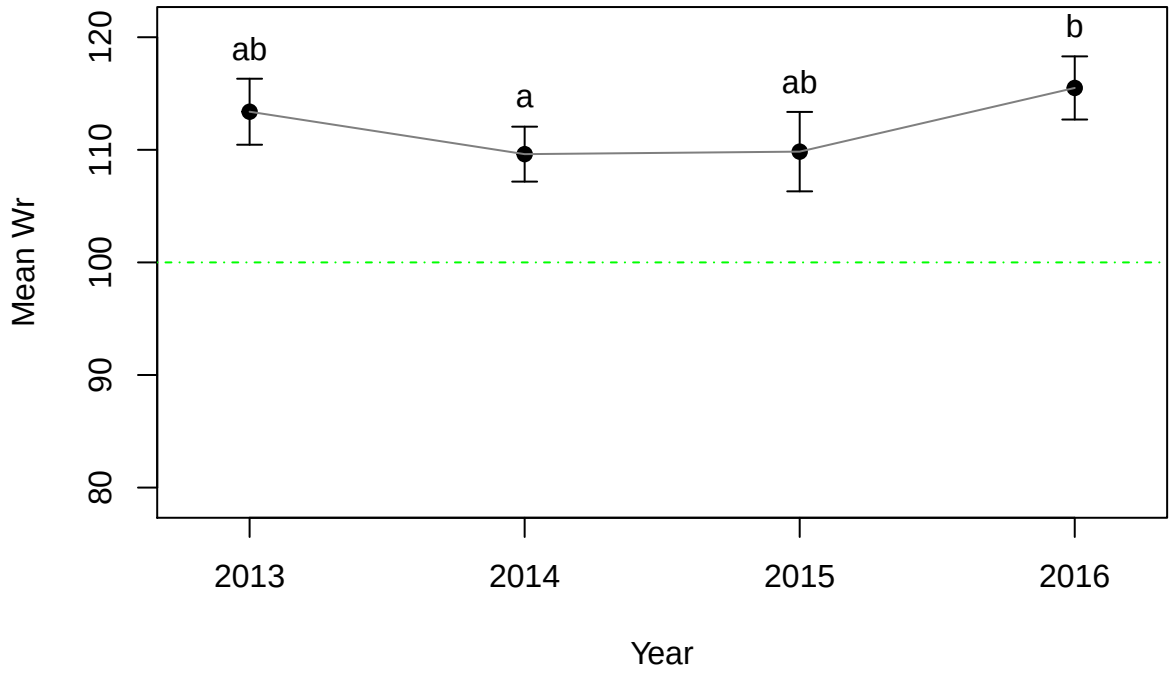
```
grps.1 <- c("2013", "2014", "2015", "2016")
nd.1 <- data.frame(Year = factor(grps.1, levels = grps.1))
(pred.1 <- predict(aov1, nd.1, interval = "confidence"))
```

```
##        fit      lwr      upr
## 1 113.3837 110.4530 116.3144
## 2 109.6148 107.1753 112.0542
## 3 109.8376 106.3113 113.3639
## 4 115.4942 112.6906 118.2977
```

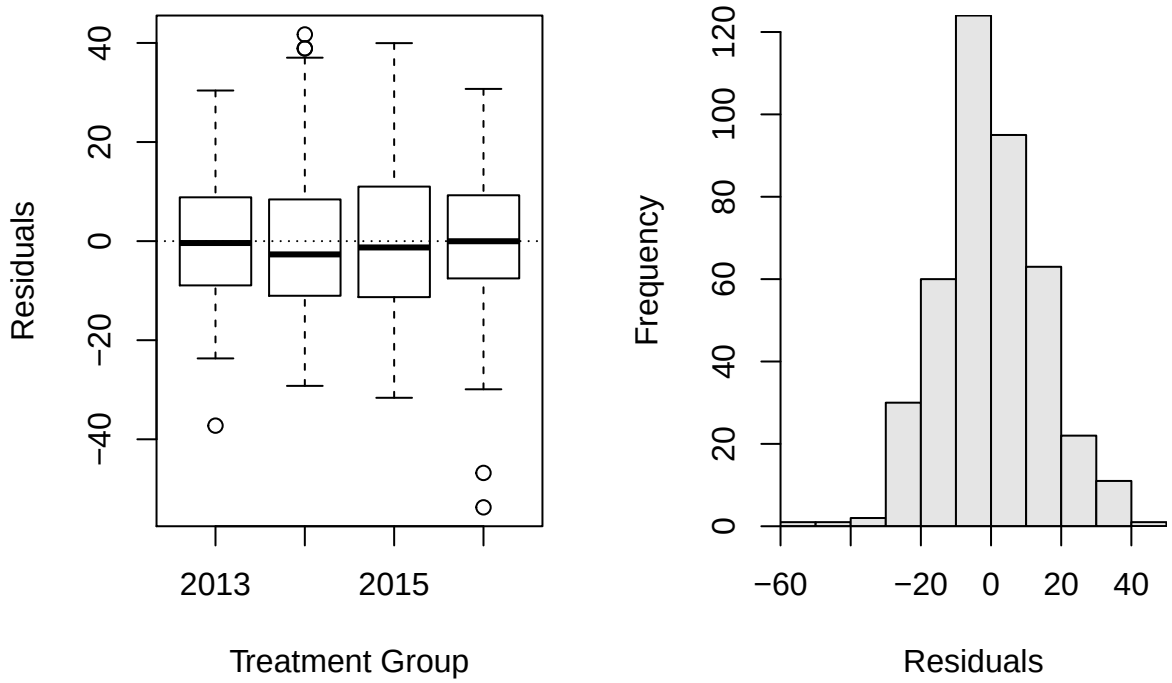
```
plotCI(as.numeric(nd.1$Year), pred.1[, "fit"], li = pred.1[, "lwr"], ui = pred.1[,
    "upr"], pch = 19, xaxt = "n", xlim = c(0.8, 4.2), ylim = c(79, 121), xlab = "Year",
    ylab = "Mean Wr")

lines(nd.1$Year, pred.1[, "fit"], col = "gray50")

axis(1, at = nd.1$Year, labels = nd.1$Year)

cld(mc1)
```

```
## 2013 2014 2015 2016
## "ab"  "a" "ab"  "b"
```

```
text(x = nd.1$Year, y = pred.1[, "upr"], labels = c("ab", "a", "ab", "b"), pos = 3)

abline(h = 100, lty = 4, col = "green")
```




```
residPlot(aov1)
```



```
leveneTest(aov1)  # Pitential outlier line 43
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   3  1.5423  0.203
##       406
```

```
# 2013 site 15 Weight = 1714 Length = 368 Wr > 200
```

Variance are equal and the homoscedasticity assumption is likley met (Levene's Test, $F_{3,406} = 1.5423$, p = 0.203).

Without 2017 *Variance are equal and the homoscedasticity assumption is likley met (Levene's Test, $F_{2,311} = 1.75$, $p = 0.18$).*

```
Year <- c("2013", "2014", "2015", "2016")
pred.1 <- data.frame(Year, pred.1)
names(pred.1) <- c("Year", "Wr", "LCI", "UCI")
str(pred.1)
```

```
## 'data.frame':    4 obs. of  4 variables:
##  $ Year: Factor w/ 4 levels "2013","2014",..: 1 2 3 4
##  $ Wr  : num  113 110 110 115
##  $ LCI : num  110 107 106 113
##  $ UCI : num  116 112 113 118
```

```
head(pred.1)
```

```
##   Year       Wr      LCI      UCI
## 1 2013 113.3837 110.4530 116.3144
## 2 2014 109.6148 107.1753 112.0542
## 3 2015 109.8376 106.3113 113.3639
## 4 2016 115.4942 112.6906 118.2977
```

```
# 2-9-2018#write.csv(pred.1,file =
# 'Data/Clean-Data/summary-data/relative-weight_largemouth-bass_STOCK.csv',row.names
# = FALSE)
```

To Be Continued...

I will look into the difference in Wr between gcat at a later date. I don't think this matters so much as of now.

```
Wr.14 <- filterD(Stock, Year == 2014)
Wr.15 <- filterD(Stock, Year == 2015)
Wr.16 <- filterD(Stock, Year == 2016)
```